DOCUMENT RESUME

ED 479 603 IR 022 207

AUTHOR	Rocha, Catarina; Xexeo, Geraldo; da Rocha, Ana Regina C.
TITLE	MyLibrary: A Web Personalized Digital Library.
PUB DATE	2002-10-00
NOTE	7p.; In: E-Learn 2002 World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education. Proceedings (7th, Montreal, Quebec, Canada, October 15-19, 2002); see IR 022 093. Figures contain illegible type.
AVAILABLE FROM	Association for the Advancement of Computing in Education (AACE), P.O. Box 3728, Norfolk, VA 23514. Tel: 757-623-7588; e-mail: info@aace.org; Web site: http://www.aace.org/DL/.
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01 Plus Postage. PC Not Available from EDRS
DESCRIPTORS	Access to Information; *Electronic Libraries; Information Retrieval; Information Services; Information Sources; Internet; User Needs (Information); World Wide Web
IDENTIFIERS	Digital Technology

ABSTRACT

With the increasing availability of information on Internet information providers, like search engines, digital libraries and online databases, it becomes more important to have personalized systems that help users to find relevant information. One type of personalization that is growing in use is recommender systems. This paper presents MyLibrary, a Web personalized digital library that provides personalized services and a recommender system (RecDoc) to support the user task of finding useful information. MyLibrary offers an environment where users can store read documents in a user profile, can rate and comment these stored documents, can visualize the documents percentage rate in a chart representation, and share information with other library users. With the information stored in the user profiles, MyLibrary offers a recommender systems, RecDoc, that indicates documents based in already read documents by the user (content-based) or based in the collaboration with other users (collaborative filtering). (Contains 13 references.) (Author)

Reproductions supplied by EDRS are the best that can be made
from the original document.

PERMISSION TO REPRODUCE AND
DISSEMINATE THIS MATERIAL IN
MICROFICHE ONLY HAS BEEN GRANTED BY

G.H. Marks

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)
2B

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
This document has been reproduced as
received from the person or organization
originating it
Minor changes have been made to
improve reproduction quality
Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy

ED 479 603

IR022207

MyLibrary: A Web personalized Digital Library

Catarina Rocha, Geraldo Xexeo, Ana Regina C. da Rocha

COPPE/UFRJ Federal University of Rio de Janeiro,
Caixa Postal: 68511 CEP: 21945-970, Rio de Janeiro - RJ, Brazil
{catarina, xexeo, darocha}@cos.ufrj.br

Abstract: With the increasing availability of information on Internet information providers, like search engines, digital libraries and online databases, it becomes more important to have personalized systems that help users to find relevant information. One type of personalization that is growing in use is recommender systems. This paper presents Mylibrary, a web personalized digital library that provides personalized services and a recommender system (RecDoc) to support the user task of finding useful information. MyLibrary offers an environment where users can store read documents in a user profile, can rate and comment these stored documents, can visualize the documents percentage rate in a chart representation, and share information with other library users. With the information stored in the user profiles, MyLibrary offers a recommender systems, RecDoc, that indicates documents based in already read documents by the user (content-based) or based in the collaboration with other users (collaborative filtering).

Introduction

The progressive growth of online information available through information services on the Internet, like search engines, digital libraries and online databases, makes people feel information overloaded. These information services provide a search and a retrieval system to the web community, but they do not provide any personalized support to individual users, or poorly support them (Amato 1999). According to (Lawrence 2000), Web search engines generally treat search requests in isolation. The results for a given query are identical, independent of the user. Consequently, few results returned by a search engine may be relevant to a user (Mizzaro 1997). Next generation search engines will make increasing use of context information from users (Lawrence 2000). One example of this attempt is the Yahoo experience in designing personalized features, presented in (Manber et al. 2000).

Providing personalized information search and delivering services, is likely to be the first step to make relevant information available to people (Amato 1999). One type of personalization that is growing in use is recommender systems.

Recommender systems improve access to relevant products and information by making personalized suggestions based on previous examples of a user likes and dislikes, or use collaborative filtering methods that built their recommendations on other users' preferences (Mooney & Roy 2000). Recommender systems have been developed in response to an actual need: helping people deal with the world of information abundance and overload. They enable people to share their opinions and benefit from each other's experience (Terveen & Hill 2001).

Digital libraries, such as others information providers, should be able to assist readers by providing personalized services and automated recommendations (Mooney & Roy 2000). Guided by this premise, a personalized environment, named MyLibrary, was developed to the Software Engineering

BEST COPY AVAILABLE

Digital Library. The library stores the scientific production of the Software Engineering group at the Federal University of Rio de Janeiro. The library also collects references to documents produced by related research centers, and universities.

MyLibrary provides a personalized environment where users can store read documents in a user profile, rate and comment these stored documents, visualize the documents percentage rate in a chart representation, share information with other library users and take advantage of their experiences to select documents evaluated with better rates. With the information stored in the user profiles, MyLibrary offers a recommender system, RecDoc, that indicates documents based on the premise that the user would like to have access to similar documents defined in his profile (documents already read by him) or documents that have been rated by similar users (collaboration with other users).

This paper is organized as follows: a first section that shows the Software Engineering Digital Library with information about the collection and the library search system. Then, a second section describes MyLibrary, a personalized environment based on user profiles that offers a recommender system. Finally summary and work in progress are shown.

Software Engineering Digital Library

The Software Engineering (SE) Digital Library centralizes the scientific production of the Software Engineering group at the Federal University of Rio de Janeiro and stores references to documents produced by others related research centers, and universities. The participants of the research group are professors, researchers, and MSc. and PhD students. In this educational environment, a library is useful to maintain an updated bibliographical collection, and to provide courses' materials. Besides, the library is available online on the Internet, so the students can search the collection to support their research activities anytime and anywhere.

According to (Entlich et al. 1997), the most important characteristic of a di gital library is its collection, whose main requirement is the selection of qualified materials. Thus, the SE group opted to maintain a manual control of documents selection and classification. They use a catalog system available at the library to register and classify documents. The library collection is composed of congress articles, journal articles, thesis, technical reports, books, and technical standards.

The SE digital library search system is based on specific fields of this type of application, such as title, author, and keyword. Besides, other search options was added to library search system to attend educational requirements of this research group, for example, search by course, by project, and by conference. All search options of the library is represented in (Figure 1). The main search interface showed in this figure is the search by keyword option where the user can combine a group of keywords or can select a keyword in the library keyword list.

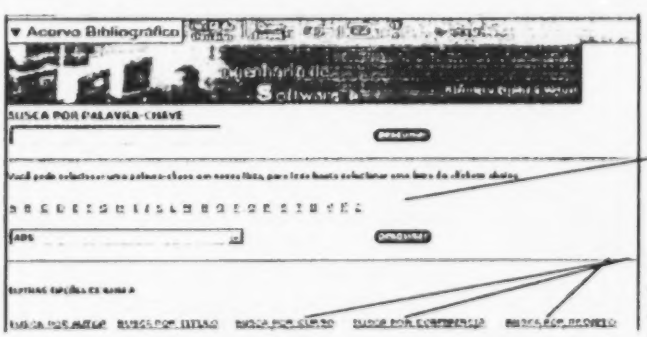

Figure 1 - Library search system screen

BEST COPY AVAILABLE

According to actual needs, digital libraries, like others information providers, should offer personalized functionalities and delivery services. Guided by this premise, a personalized environment (MyLibrary) and a recommerder system was developed to the SE Digital Library. MyLibrary is described in the following section.

MyLibrary: a personalized environment to the SE digital library

Concerned to provide personalized services to the SE Digital Library, the first step toward personalization was the addition of a user profile to represent user information needs. The second step was the development of a recommender system that manipulates the user profiles to find relevant information. The resulting personalized environment, MyLibrary, is available only to registered users. The library search system is still available to any user that accesses the library, as ilustrated in (Fig. 2). This figure also shows MyLibrary architecture.

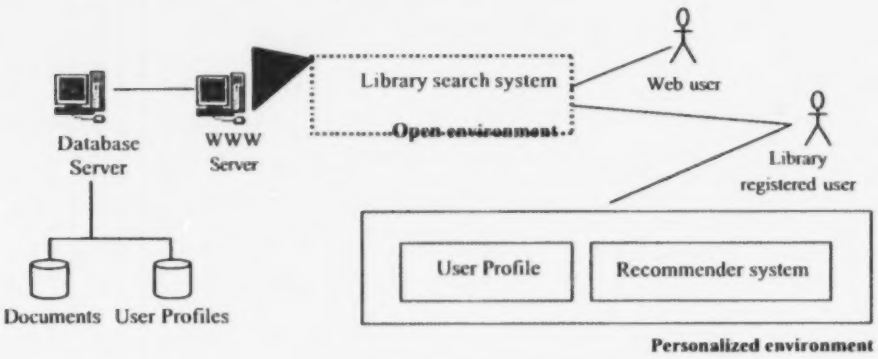

Figure 2 - MyLibrary architecture

The *Web user*, not registered at the library, has access to the environment called "open system". Through the open system, the user interacts with the library search system, querying documents by title, author, keyword, course, project, and conference.

The *DL registered user* has access to the personalized environment and the recommender system. There are two databases composing the library: Documents database that stores all documents; Profiles database with the user profiles representation.

MyLibrary embraces personalized services, such as keep document in the user profile, rate and comment these stored documents, and send a document to a friend by e-mail, as shown in (Fig. 3). This figure also represents the document description, average rate, a graphical representation of all rates for the document, and a list of comments made by the users.

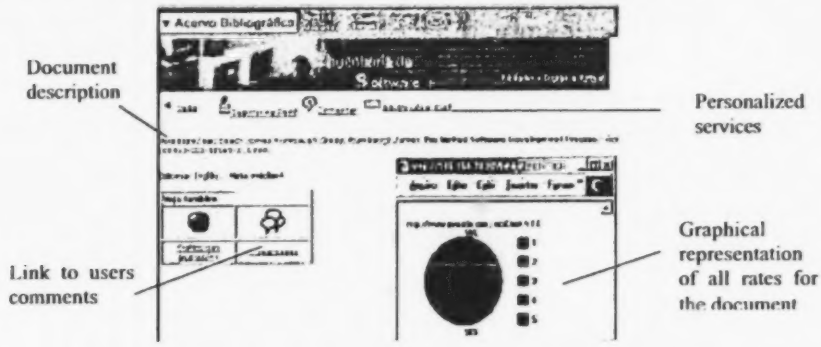
Document description
Personalized services
Link to users comments
Graphical representation of all rates for the document

Figure 3 – Document description and personalized services at MyLibrary

BEST COPY AVAILABLE

With the documents evaluation, a average rate can be associated to documents and displayed in the list of documents returned by the library search system. A user can takes advantage of others users experience to choose better rated documents in this list. Besides, the users' comments are showed when the document is selected in list, as illustrated in (Fig. 3).

User profile

A user profile schema to digital libraries, proposed in (Amato 1999), was adapted to MyLibrary according to particularities of the SE Digital Library. The resulting profile schema is showed in (Tab. 1) below, and includes all information to be represented in the user profile.

Personal data category	• Name, e-mail, password and user occupation (student, professor, etc.)
Gathering data category	• A user keyword list that represents user interesting area. • Digital documents are available in PDF format. • The user only query the library database.
Delivering data category	• Delivery mode: the system sends an e-mail to the user suggesting new interesting documents. • Delivery time: the recommended items is available in the personalized environment when the user enters the system.
Actions data category	• Stored read documents. • Rated documents (from 1 to 5) • User relevance feedback (the users explicit rate documents and recommendations).

Table 1 - Mylibrary user profile schema

Initially, the profile is created when the user submits a registration form available at the library. This form contains information about personal data, academic data, and a list of keywords. While the user interacts with the library search system, stores documents in the user profile, submits rates and comments about them, the profile has been continually updated. The list of read documents and their evaluations are available at the user profile, as illustrated in (Fig. 4).

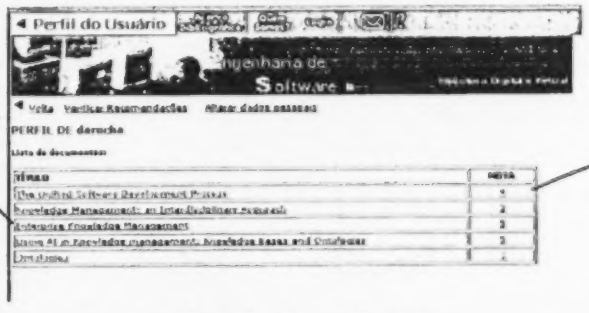

Figure 4 - User profile screen

With the information stored in the user profiles, a delivery service has been implemented through a recommender system named RecDoc. The system is described in the section below.

BEST COPY AVAILABLE

RecDoc: Document Recommender System

Having emerged in response to technological possibilities and human needs created by the WWW, recommender systems aim to mediate, support and automate everyday process of sharing recommendations (Terveen & Hill 2001). There are two main recommendation systems approaches (Balabanovic & Shoham 1997): content-based and collaborative filtering. In the content-based approach, the document recommendation is done through comparison of similar documents to those the user has already seen and rated . The collaborative approach explores techniques for matching people with similar interests and then making recommendations on this basis (Sarwar et al. 2001).

Content-based approach has its origins in the Information Retrieval research community and employs many of IR techniques and algorithms. Content-based recommenders believe in the intuition "find me things I have liked in the past". Documents are recommended based on the comparison between their content and user profiles (Balabanovic & Shoham 1997). Systems that use this approach are InfoFinder and NewsWeeder (Lang 1997).

Recommendation systems adopt collaborative filtering based on the premise that people looking for information would like to know what other people already found and evaluated. Systems that use this approach are GroupLens (Resnick et al. 1994), Fab (Balabanovic & Shoham 1997) and Firefly .

RecDoc uses hybrid approach taking advantage of these two approaches: content-based and collaborative. The Fab system, developed by Stanford University, also uses the hybrid approach for recommendation of Web pages.

RecDoc provides to the user:

1. to receive, for each database update, new documents that are interesting to the user trough the comparison between the document keywords and the user profile **(content-based approach)**.
2. to receive documents rated by similar users **(collaboration approach)**.

RecDoc applied one K-nearest neighbor algorithm to perform the collaborative filtering recommendation. These algorithm is one of the most successful recommendation technique to date on the Web (Sarwar et al. 2001).

Every time that RecDoc indicates a document, the user is suggested to evaluate the recommendation, accomplishing what is called relevance feedback. Accurate profiles are so much important in the content approach (to guarantee appropriate recommendations) as in the collaboration approach (to guarantee that users with similar profiles are really similar).

Summary and Work in progress

This paper presented the SE Digital Library, available on the Internet. The library has been in use to support activities and courses of the software engineering research group at the Federal University of Rio de Janeiro.

Then, MyLibrary a personalized environment based on user profiles and a recommender system (RecDoc) was described. RecDoc is an implementation of a hybrid content-based/collaborative document recommender system. The recommendation of documents seems to be a useful service to deliver relevant information to the library users.

At this moment, a prototype of RecDoc has been evaluated trough a testbed database.

References

1. Amato, G., Straccia, U. (1999). User Profile Modeling and Applications to Digital Libraries. Proceeding of ECDL Conference, Paris.
2. Lawrence, S. (2000). Context in Web Search. IEEE Data engineering Bulletin, 23(3): 25-32.
3. Mizzaro, S. (1997). Relevance: the whole history. Journal of American Society for Information Science, 48(9):810-832.
4. Manber, U., et al. (2000). Experience with Personalization on Yahoo!. Communications of ACM, volume 43(8): 35-39.
5. Terveen, L., Hill, W. (2001). Beyond Recommender System: Helping People to Find Each Other. Proceedings of HCI in the New Millennium, Jack Caroll, ed., Addison-Wesley.
6. Mooney, R., Roy, L. (2000). Content-based Book Recommending Using Learning for Text Categorization. Proceedings of ACM Digital Libraries, pp.195-204.
7. Entlich, R. et al (1997). Making a digital library: the contents of the CORE project. Communications of ACM, 15(2).
8. Resnick, P., Varian, H. (1997). Recommender Systems. Communications of ACM, 40(3).
9. Resnick, P. et al (1994). GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In proceedings of ACM Conference on Computer-Supported Cooperative Work.
10. (Balabanovic & Shoham 1997) Balabanovic, M., & Shoham, Y., Fab: Content-Based, Collaborative Recommendation. Communications of the ACM, 40(3): 66-72.
11. Sarwar, B., et al. (2001). Item-based Collaborative Filtering Recommendation Algorithms. In: proceedings of WWW10 Conference.
12. Lang, K. (1997). NewsWeeder: Learning to Filter NetNews. In Machine Learning: Proceedings of the Twelfth International Conference. Lake Tahoe, California: Morgan Kaufmann.
13. Shardand, U., & Maes, P. (1995). Social Information Filtering: Algorithms for Automating "Word of Mouth". In Proceedings of the CHI-95 Conference, Colorado.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

[X] This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

[ ] This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").